DOCUMENT RESUME

CG 007 511

24

ED 064 662

AUTHOR Klausmeier, Herbert J.; And Others
TITLE Research and Development Strategies in Theory Refinement and Educational Improvement.
INSTITUTION Southwest Regional Educational Lab., Inglewood, Calif.; Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
REPORT NO TP-15
BUREAU NO BR-5-0216
PUB DATE Nov 68
CONTRACT OEC-5-10-154
NOTE 30p.; Papers presented at the symposium of the American Educational Research Association, Chicago, Illinois, February, 1968

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Cognitive Objectives; *Educational Improvement; *Educational Research; *Instructional Materials; *Learning Theories; Research and Development Centers

ABSTRACT Research and development strategies designed and implemented at the Wisconsin Research and Development Center for Cognitive Learning and the Southwest Regional Laboratory are discussed in papers by the directors of each and by members of the R & D Center staff. These papers further give examples of the research and development activities underway in each organization. Papers on the controlled experiment focus upon its applicability in different settings. The large-scale experiment involving many classes uses the class as the experimental unit, whereas smaller experiments conducted in the newly devised Research and Instructional Units may use the individual as the unit. The steps in developing and refining a prototypic instructional system are clearly outlined in another paper. Finally, the usefulness of formative evaluation procedures in effecting educational improvements are discussed. (Author)

ED 064662

THEORETICAL PAPER NO. 15

RESEARCH AND DEVELOPMENT STRATEGIES IN THEORY REFINEMENT AND EDUCATIONAL IMPROVEMENT

JOINT PUBLICATION
TECHNICAL
RESEARCH

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

WISCONSIN RESEARCH AND DEVELOPMENT

ER FOR
NITIVE LEARNING

CG 007 511

BR 5-0216
PA 24
CG

CENTER FILE COPY

ED 064662

Theoretical Paper No. 15

RESEARCH AND DEVELOPMENT STRATEGIES IN THEORY REFINEMENT AND EDUCATIONAL IMPROVEMENT

By Herbert J. Klausmeier, James L. Wardrop, Mary R. Quilling, Thomas A. Romberg, and Richard E. Schutz

With remarks by Leslie D. McLean

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

and

The Southwest Regional Laboratory
for Educational Research and Development
Inglewood, California

November 1968

These papers were initially presented at a symposium of the American Educational Research Association, Chicago, Illinois, February, 1968. The preparation of the papers and the research reported herein were performed pursuant to contracts with the United States Office of Education, Department of Health, Education, and Welfare, under the provisions of the Cooperative Research Program.

PREFACE

The first systematic attempts to apply research and development strategies to education occurred with the establishment of research and development centers in 1964 and regional educational laboratories in 1965. However, these strategies as they applied to education were neither clearly defined nor well tested. It has become the task of R & D centers as well as regional laboratories to refine existing strategies on the basis of experience and to invent new ones.

It is appropriate, then, that the Wisconsin Research and Development Center for Cognitive Learning and the Southwest Regional Laboratory for Educational Research and Development cooperate in presenting strategies currently employed at each institution to the scholarly community in a symposium at the American Educational Research Association meetings. Professor Klausmeier, organizer and chairman of the symposium, outlines R & D strategies of the Wisconsin R & D Center in his introduction. Two applications of controlled experimentation in school settings are described by Dr. Wardrop and Mrs. Quilling. Dr. Romberg presents the procedure to be utilized in the development and refinement of prototypic instructional systems. Representing the regional laboratory approach is Dr. Schutz, who considers strategies for improving instruction through product development. Finally, Professor McLean reacts to the strategies presented, comparing them with those used at a Canadian counterpart to R & D centers.

James E. Walter
Director of Dissemination

CONTENTS

		Page
	List of Tables and Figures	vi
	Abstract	vii
I.	Research and Development Strategies in Education Herbert J. Klausmeier	1
II.	Controlled Experimentation in Multiclassroom Settings James L. Wardrop	6
III.	Controlled Experimentation in Research and Instruction Units Mary R. Quilling	10
IV.	The Development and Refinement of Prototypic Instructional Systems Thomas A. Romberg	14
V.	Experimentation Relating to Formative Evaluation Richard E. Schutz	19
VI.	Remarks Following the Presentation of Papers Leslie D. McLean	23

LIST OF TABLES AND FIGURES

		Page
Table		
1	Means on Dependent Variables for Assessing the Effect of Feedback on Acquisition of Mathematics Concepts	11
2	Means on Teacher-Constructed Test to Assess the Effect of Models on Learning Mathematics Concepts	12
3	Analyses of Variance on Scores of Teacher-Constructed Test	12

Figure		
1	Relationships among Research and Development Activities at the Wisconsin Research and Development Center for Cognitive Learning	2
2	Steps in Research on Learning—Pure Research to Technological Development	4
3	A Classification Schema of Processes Related to and Necessary for Change in Education	5
4	Model of an Instructional System	15

ABSTRACT

Research and development strategies designed and implemented at the Wisconsin Research and Development Center for Cognitive Learning and the Southwest Regional Laboratory are discussed in papers by the directors of each and by members of the R & D Center staff. These papers further give examples of the research and development activities underway in each organization.

Papers on the controlled experiment focus upon its applicability in different settings. The large-scale experiment involving many classes uses the class as the experimental unit, whereas smaller experiments conducted in the newly devised Research and Instructional Units may use the individual as the unit. The steps in developing and refining a prototypic instructional system are clearly outlined in another paper. Finally, the usefulness of formative evaluation procedures in effecting educational improvements are discussed.

I

RESEARCH AND DEVELOPMENT STRATEGIES IN EDUCATION

Herbert J. Klausmeier
Professor of Educational Psychology
The University of Wisconsin
Director, Wisconsin R & D Center

The purpose of this collection of papers, as of the symposium from which it developed, is to identify and discuss contributions that can be made simultaneously to improving instruction and to refining teaching-learning theory through the use of relevant research and development strategies applied to education. This kind of research and development is of very recent origin. The first attempt to apply research and development strategies systematically to education was initiated by the U.S.O.E. in 1964 when research and development (R & D) centers, including the Wisconsin Research and Development Center for Cognitive Learning, were established on four university campuses. As of January 1968, there were nine R & D centers. Each center has identified a problem area—such as cognitive learning, educational stimulation in early childhood, evaluation, or teacher behavior—as its focus. Each center conducts research in order to extend knowledge and theory related to the problem area and also engages in related development to improve educational practices.

In 1965 the Elementary and Secondary Education Act was passed. In connection with this legislation, 20 regional educational laboratories had been established as of January 1968. The regional laboratories are incorporated as separate legal entities and are not affiliated with institutions of higher learning. In general, a regional laboratory engages quite heavily in development and development-related research while an R & D center has a somewhat greater emphasis on basic research and a lesser emphasis on development.

At present then there are 9 R & D centers and 20 regional laboratories engaged in research and development designed to improve educational practices. Different from long established and well funded research and development in agriculture, space, and industry, research and development in education does not have clearly defined and well tested strategies. So the staff of the Wisconsin R & D Center is engaged not only in attempting to extend knowledge about cognitive learning and to improve related educational practices, but also in inventing research and development strategies in education and refining existing ones.

Before considering some detailed strategies employed in the Wisconsin R & D Center for Cognitive Learning, its focus and global strategies should be considered. In the R & D Center we differentiate between basic and applied research in terms of purposes. Basic research is conducted to generate knowledge and is not concerned with whether the knowledge may be useful in improving educational practice. Development-based research, referred to by some as applied research, is specifically concerned with developing a substantive or procedural output designed to achieve specified objectives and with ascertaining how well it achieves the objectives under varying conditions with children or instructional personnel of varying characteristics.

The prevalent form of basic research is the controlled experiment and its variants. Most controlled experiments designed to extend knowledge about learning are conducted in the laboratory, or under laboratory conditions. Other types of basic research employed in education include correlational, factor analytic, survey, and case study. These studies are often conducted in the school setting, with children, school personnel, or both as subjects.

The two distinguishing features of development-based research are its purposes, as given

before, and the procedures employed. Development-based research encompasses identifying a deficiency or problem area in some component of an instructional system; identifying or formulating objectives of that component; and developing, testing, and refining a new component to achieve the objectives. Development-based research may be initiated by practitioners or by scholars. The need for it is experienced by practitioners directly and by researchers through formal and informal surveys. Small developmental projects may be initiated and executed by school personnel. Comprehensive projects designed to produce an instructional system in a subject field at the elementary school level may run for five or more years. The testing and refinement of a substantive or procedural output through various revisions may involve controlled experimentation, correlational studies, and case studies. It must involve ascertaining how well the output achieves the objectives for which it was developed.

The evaluation of instructional materials, procedures, and equipment that have not been tested in the school system may be designated as a third kind of research. The purpose here is to improve instruction, not to add to knowledge. Again, several forms of research may be executed in an evaluative study; however, answers must be sought concerning how well the product performs in connection with clearly stated criteria.

Figure 1 indicates that basic research and developmental research are conducted in each of the two major programs of the Center. In Program 1 basic research is conducted to extend knowledge about processes and conditions of learning, whereas in Program 2 the basic research is conducted to extend knowledge about instructional processes. The straight line connecting the two sets of basic research activities indicates the relationship of similar methods and concepts in the two programs. However, many variables not considered in learning research must be taken into account in research on instruction, particularly those variables associated with teacher behaviors and characteristics, student behaviors and characteristics, subject-matter content and sequence, utilization of time, utilization of space, and organization of personnel for instruction.

The other relations between basic research and development-based research shown in Figure 1 are of primary interest. Obviously, there

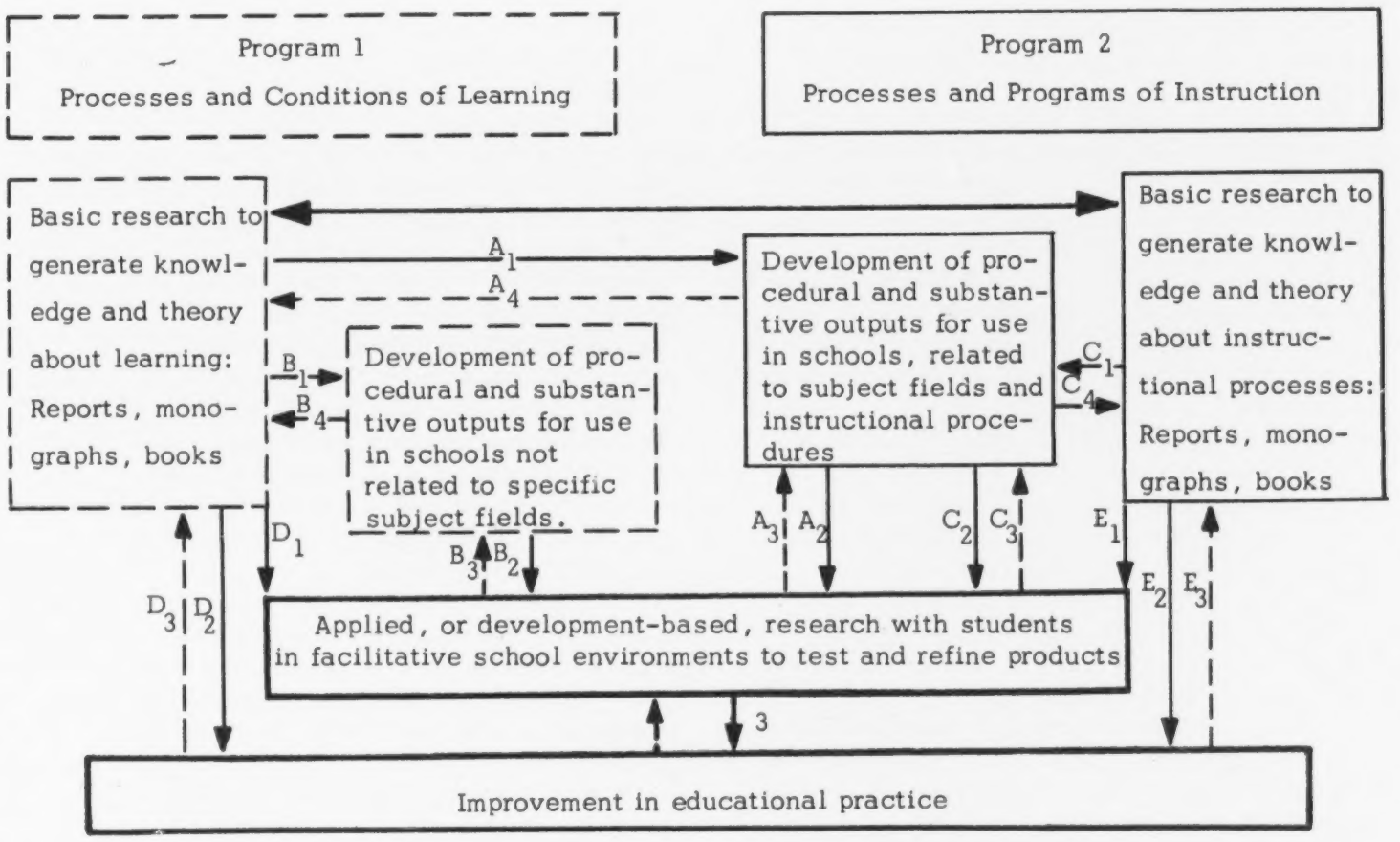

Figure 1. Relationships among research and development activities at the Wisconsin Research and Development Center for Cognitive Learning

are many relations, rather than one linear relationship. In the $A_1 \rightarrow A_2 \rightarrow 3$ sequence, basic research in the behavioral sciences provides the foundation for the development of procedureal and substantive outputs, such as instructional materials and equipment for use by students or descriptions of procedures to be used by teachers and other educational workers. In turn, these outputs are researched and refined with school-age children in laboratory situations or directly in school settings. In actual practice in our R & D Center there is not a clear-cut terminal point at which research results are stated and product development begins, nor are there separate groups of personnel, one doing research and another product development. Rather, the combined research and development strategy here involves cooperative effort by specialists in learning, the subject disciplines, and methodology and also teachers or other school personnel in Multiunit Schools.

The $B_1 \rightarrow B_2 \rightarrow 3$ and the $C_1 \rightarrow C_2 \rightarrow 3$ sequences are more typical of present research and development activities at the Wisconsin R & D Center. Here one or more investigators with similar interests in either learning or instruction do the basic research, develop a substantive or procedural outcome intended for students or teachers, and carry through with the research in school settings. Consultants with various specialties and school personnel participate in the activities at appropriate points.

So that specialists in various fields may be available to several projects, regardless of size and personnel combinations, the Wisconsin R & D Center is developing a large technical section, staffed by persons of varying specialties who participate simultaneously in several projects of both programs. Also, in order to secure continuous cooperation with school personnel, Multiunit Schools that facilitate research and development activities have been organized.

The three preceding sequential strategies are operative in the Wisconsin R & D Center for Cognitive Learning. There are other sequences. Lines D_1 and E_1 indicate that, although some basic research may terminate in relevant reports, monographs, or books, the results may lead directly to developmental research in school settings, without the intermediate step of developing a product to be used by children or teachers. Also, lines D_2 and E_2 show that some research results may be put directly to use in educational improvement. Both sequences occur in our Multiunit Schools when teachers participate in experiments, secure feedback about the results, and put the relevant results into practice.

The preceding discussion has focused on ideas and activities moving in the general direction of research → development → practice. As shown in Figure 1 by the dotted lines, the sequence may be in the other direction. For example, the desire on the part of school people and others to improve the education of disadvantaged children has led to the formulation of more precise educational specifications for the subsequent development of substantive and procedural products. The same practical problem has also generated some basic research dealing with cognitive skills. The continuous interaction of the Wisconsin R & D Center staff with school people, and also the continuous examination of educational problems by the R & D staff, have markedly influenced the development activities of the Center. At the present time much basic research has been generated in connection with developmental activities. Probably a lesser amount of basic research has led directly to development.

The actual sequencing of activities is not precise in either direction. Rather, there is more a continuous interplay of researching and developing. The personnel primarily engaged in improving an instructional system conduct comprehensive development-based research to ascertain how well the program works with the clients or consumers and make changes accordingly. Although not conducting basic research, they refer to contemporary theory and also to the results of basic research through reading and attending conferences and by bringing in knowledgeable consultants. Thus, a community of scholars representing various specialities, some within the Center on a more permanent basis and others from outside the Center on a consulting basis, is an essential component of an effective research and development strategy in education.

In the preceding discussion a number of important points have been made. First, the initiation of activities designed to improve educational practice may be at any of three points: an analysis of deficiencies in educational practice, the design and development of instructional products, or the conduct of basic research. Second, the output from basic research is knowledge, while from development and applied research the outputs are instructional materials, procedures, organizations, etc. that are designed specifically to improve educational practice. Third, there are several relations among research, development, and improved practice rather than a single linear research → development → practice sequence.

The latter needs special emphasis in view of two widely disseminated models of research

→ development → practice. In 1964 Hilgard outlined a viewpoint of many educational psychologists as shown in Figure 2. In this sequence it is assumed that learning theory has relevance for improving instruction but that development activities and applied research are essential intervening steps. Hilgard points out that the six steps are not discrete, but are on a continuum, and that invention plays an important role in bringing about improved practice. Nevertheless, there is a strong tendency toward emphasizing basic research as both the inititation point for and primary contributor to subsequent development and thence improved practice.

The Guba-Clark model is shown in Figure 3. This model also puts research, development, diffusion, and adoption in a linear arrangement from left to right. Guba and Clark noted that the model was a unidimensional analysis of change roles influenced by many variables and also that change did not necessarily begin with research. While Guba and Clark may accept the kind of relations outlined in the model of the Wisconsin R & D Center, they did not make the same relations explicit in their model. Furthermore, there is a widespread tendency, particularly among behavioral scientists interested in education, to assume that the primary basis for improved educational practice is the results of basic research.

More specific strategies in research and development are presented next. The first three speakers in this symposium draw from their experiences at the Wisconsin Research and Development Center for Cognitive Learning. Richard E. Schutz, Director of the Southwest Regional Laboratory for Educational Research and Development at Lo- Angeles, presents their strategies for developing and evaluating instructional materials. Leslie D. McLean of the Ontario Institute for Studies in Education discusses the strategies and techniques presented. Let me point out that in this symposium we are not engaged in painting pictures of something that may emerge in the decades ahead. The Wisconsin R & D Center and the Southwest Regional Educational Laboratory may be visited, and our strategies may be observed and discussed. We are, however, vitally concerned with the future and are thus open to suggestions that may help us in developing more creative research and development strategies to improve education.

PURE RESEARCH			TECHNOLOGICAL RESEARCH AND DEVELOPMENT		
Not Directly Relevant	Relevant Subjects and/or Topics	School-relevant Subjects and Topics	Laboratory, Classroom, and Special Teacher	Tryout in "Normal" Classroom	Advocacy and Adoption
Step 1	Step 2	Step 3	Step 4	Step 5	Step 6
Animal Mazes, Eyclid Conditioning, Pursuit Learning, etc.	Human Verbal Learning, Concept Formation, etc.	Mathematics, Reading, Typing, etc.	Programed Instruction; Language Laboratory, in Early Stages	Results of Step 4 Tried in Regular Setting	Manuals and Textbooks Prepared; Teacher Training Undertaken

Figure 2. Steps in research on learning— pure research to technological development.

Source: Hilgard, Ernest R. A perspective on the relationship between learning theory and educational practice. In Hilgard, Ernest R. (Ed.) Theories of Learning and Instruction. Thirty-sixth Yearbook. Chicago: National Society for the Study of Education, 1964.

	Research	Development		Diffusion		Adoption		
		Invention	Design	Dissemination	Demonstration	Trial	Installation	Institutionalization
Objective	To advance knowledge	To formulate a new solution to an operating problem or to a class of operating problems, i.e., to innovate	To order and to systematize the components of the invented solution; to construct an innovation package for institutional use, i.e., to engineer	To create widespread awareness of the invention among practitioners, i.e., to inform	To afford an opportunity to examine and assess operating qualities of the invention, i.e., to build conviction	To build familiarity with the invention and provide a basis for assessing the quality, value, fit, and utility of the invention in a particular institution, i.e., to test	To fit the characteristics of the invention to the characteristics of the adopting institution, i.e., to operationalize	To assimilate the invention as an integral and accepted component of the system, i.e., to establish
Criteria	Validity (Internal and external)	Face Validity (appropriateness) --- Estimated Viability --- Impact (relative contribution)	Institutional Feasibility --- Generalizability --- Performance	Intelligibility --- Fidelity --- Pervasiveness --- Impact (extent to which it affects key targets)	Credibility --- Convenience --- Evidential Assessment	Adaptability --- Feasibility --- Action	Effectiveness --- Efficiency ---	Continuity --- Valuation --- Support
Relation to Change	Provides basis for invention	Produces the invention	Engineers and packages the invention	Informs about the invention	Builds conviction about the invention	Tries out the invention in the context of a particular situation	Operationalizes the invention for use in a specific institution	Establishes the invention as a part of an ongoing program; converts it to a "non-innovation"

Figure 3. A Classification Schema of Processes Related to and Necessary for Change in Education

Source: Guba, E. G., & Clark, D. L. An examination of potential change roles in education. Paper read at Symposium on Innovation in Planning School Curricula sponsored by the National Education Association Committee for Study of Instruction, Airielhouse, Va., October, 1965, P. 8.

II

CONTROLLED EXPERIMENTATION IN MULTICLASSROOM SETTINGS

James L. Wardrop[1]
Assistant Professor of Educational Psychology
Center for Instructional Research and Curriculum Evaluation
University of Illinois

"Research" in education has many faces. It may appear as an associational study, a case study, an informal description of events, a controlled experiment, or any number of other activities. For the purpose of elucidating causal relationships—answering questions of the form, "What is the effect of variable x on behavior Y?" (where X and Y are probably both multivariate rather than univariate)—the appropriate face (or should I say mask?) is the controlled experiment.

The phrase "controlled experiment" carries with it some obvious connotations: there is some independent variable or variables which the experimenter manipulates; there are two or more groups which are exposed to different levels of these independent variables; there may be an attempt to make these groups homogeneous with respect to characteristics, like IQ and past achievement, which are not under the experimenter's control; there is some dependent variable which is amenable to being measured with sufficient precision that differences may be detected when they exist; this criterion variable is one which is expected to be affected by the experimental manipulations; and the criterion scores are subjected to some sort of analysis in order to arrive at some decisions about the effectiveness of the experimental manipulations.

In addition to these rather basic characteristics, there are some other very important aspects to controlled experimentation. Consider, for example, the objective of eliminating as many as possible of the alternative explanations that might reasonably be offered for the results of an experiment. Campbell and Stanley (1963) have presented an excellent discussion of some of these alternatives— factors which might bias the outcome of an experiment such as selection, maturation, and statistical regression. For a thorough discussion of such factors, I refer you to the Campbell-Stanley monograph. For a good nonstatistical discussion of experimental design, David Cox's Planning of Experiments (1958) has much to recommend it.

In educational research, as in many other scientific enterprises, the argument over basic versus applied research continues to rage. (See, e.g., Cronbach, 1966, and Ebel, 1967.) How applicable, for example, are the results of laboratory experiments to the natural setting of the classroom? Can the results of basic research be applied on a large scale in classroom settings? I am now raising questions about the generalizability of results, what Campbell and Stanley call "external validity." It is at this point that the need for controlled experiments in multiclassroom settings becomes apparent.

Much of the research that has been done in education has been justly criticized because statistical analyses have been based on individual students' performance when the appropriate experimental unit was the classroom. (See Cox, 1958, p. 2; Lindquist, 1940, pp. 21-24.) In this connection, it is worth quoting Lindquist, who said:

> In general...many of the samples employed in educational research consist of a small number of intact groups (such as classes in the same or different schools, groups of pupils in separate buildings in the same system, or

[1] During 1966-1967 Dr. Wardrop was a Title IV postdoctoral fellow at the Wisconsin Research and Development Center for Cognitive Learning.

> groups of pupils from different communities or geographical regions), or of a small number of subsamples selected from different "strata" in the population....In all such cases, the "size" of the sample is dependent, not upon the number of individual observations, but upon the number of intact groups or subsamples of which the total sample is constituted. In other words, the unit of sampling in educational research is often the class, the school, or the community, rather than the pupil.

Julian Stanley (1965) made explicit the fact that

> In determining what the experimental unit is, it makes no difference whether pupils were assigned at random to the classrooms or systematically. The basic point is that the tth treatment was assigned at random to the classroom as a whole, rather than just to certain individuals within it randomly.

The point of all this is as follows: if one wishes to determine the effectiveness of some curriculum package (which is administered by a teacher to a class), or of some teaching method (again administered by a teacher to a class), or of group size (where instruction is given to small or large groups), or of some scheduling system (where, for example, groups of students are assigned either modular or traditional class schedules), one is faced with the situation in which the experimental unit, and hence the unit of analysis, is a group or class of students, not the individual student.

In attempting to determine the applicability of laboratory findings in classroom settings, it is imperative for internal validity that controlled experiments be used, and for both internal and external validity that these experiments be carried out in a number of classrooms or schools and the classroom or school be treated as the unit of analysis.

To illustrate the kinds of problems being investigated through this particular approach, let us consider some specific multiclassroom experiments being conducted by the Wisconsin R & D Center. One project, "Situational Variables and Efficiency of Cognitive Learning," is designed to investigate selected variables and conditions associated with efficient learning of concepts and cognitive skills and to develop and test a system to facilitate concept learning, a system of motivation, and a system of individualization in school settings. As a direct outgrowth of an earlier project designed to clarify functional relationships among variables related to efficiency of cognitive learning in laboratory settings, this project represents a direct attempt to validate laboratory findings in the classroom, thus extending knowledge about theoretical relationships and at the same time improving classroom instruction.

Another project which is using multiclassroom experimentation deals with English syntax and composition. This project has provided materials for experimentation on learning variables in school settings (Blount, Klausmeier, Johnson, Fredrick, & Ramsay, 1967) and developed instructional materials for use by teachers and students in junior high schools. Here again the project has the two purposes of generating and refining knowledge about variables affecting classroom learning and of improving instruction in English grammar and composition. While the major emphasis in the previous project was on theory refinement, the main focus of this project is on improving classroom instruction.

In order to illustrate something of the nature of multiclassroom experimentation, let me describe briefly one recent project in which I was involved. The experiment was carried out in the Racine, Wisconsin, Unified School District under the joint sponsorship of the Wisconsin R & D Center and the Creative Thinking Project at the University of California in Berkeley (Olton, Wardrop, Covington, Goodwin, Crutchfield, Klausmeier, & Ronda, 1967). Its major purpose was to investigate the extent to which increments in the thinking and problem-solving performance of fifth-grade students could be brought about by the use of programed lessons (Covington, Crutchfield, & Davies, 1966) which were designed to teach skills and strategies of creative thinking independent of any specific subject field. Of the 47 fifth-grade classrooms in the school district, 44 were used in this experiment, 3 classrooms having been randomly eliminated. For all classrooms, ratings were obtained of the extent to which the overall classroom atmosphere seemed to provide an environment that facilitated creative thinking on the part of the students. Data were also obtained on the IQs and achievement levels of all students in these classes. The 44 classes were then divided into 22 matched pairs such that both classes in each pair had virtually the same environment rating and were also similar with respect to mean IQs and achievement test scores. One class from each pair was then randomly assigned to the experimental group (which would use The

Productive Thinking Program). The other class from each pair then was placed in the control group. All students were given a battery of tests of creative problem solving to provide baseline information. At the conclusion of the experimental training program, another similar battery of tests was administered. After the posttests were administered but before any scoring was done, eight males and eight females were selected from each class as follows: within each class, all male students were rank-ordered on the basis of IQ and one student randomly selected from each eighth of the distribution. Eight females were then selected from each class in the same way, so that the statistical analysis was based on the scores of 44 x 16 = 704 students.

However, the analysis was carried out using mean scores for males and females in each class. We thus had a design with three factors (treatment group, classroom environment, and sex) and 11 observations (classrooms) per cell, not the 88 per cell we would have had using individual student scores as the unit of analysis. The consequence of this should be obvious: the within-cell variance estimate in the analyses of variance had 80 degrees of freedom associated with it instead of 696. In this experiment, this was not an important limitation. (An F ratio of 3.96 is required for significance at the .05 level with 1 and 80 degrees of freedom, as contrasted with an F of 3.86 with 1 and 696 d.f.) It is easy, however, to conceive of situations in which the loss of power because of the limited number of degrees of freedom would be quite severe.

To return to the creative-thinking experiment, I would like to indicate briefly the contributions of this particular project to theory refinement. The results provide evidence that the training materials are effective in developing creativity and problem-solving skills. In conjunction with other research, they indicate the necessity for teacher participation as an integral part of the training program. An important question must be asked in view of the findings concerning IQ. There was a significant relationship found between IQ and performance on the creativity and problem-solving tasks used, but no significant IQ by treatment interactions. It is surprising to find these relationships. One could readily justify expecting just the opposite relationship (that is, an IQ by treatment interaction but perhaps not a significant relationship between IQ and performance on the creativity tasks).

The experiment just described is what Campbell and Stanley (1963) call design No. 4, the Pretest-Posttest Control Group design. As such, it has high internal validity (freedom from bias). However, because of such possible limitations to external validity as what Campbell and Stanley have called the interaction of testing and the experimental treatments (partially controlled by using some tests very unlike anything in the training materials), and particularly because the experimental sample was taken from a population which was limited in some ways but was in many ways representative of students with a wide range of backgrounds and abilities, any generalizations beyond the Racine Unified School District must be made with caution. In the May 1967 American Educational Research Journal, Hickrod described what he called a "metropolitan case study." Although the problem-solving research in Racine was a multiclassroom experiment, it is in its sampling and generalizability aspects much like Hickrod's metropolitan case study.

In this presentation, I have attempted to indicate the value of controlled experimentation in multiclassroom settings. I have not yet considered the kinds of resources such research demands. But resources are an important consideration. Multiclassroom experimentation places greater demands on the staff and requires stronger financial support than does most educational research. It is for this reason that it is most likely to develop in the context of large-scale activities, such as the major curriculum development projects, research and development centers, and perhaps the regional laboratories.

Most importantly, though, such research requires the wholehearted cooperation of the schools. They must be willing to commit the time of administrators, teachers, and students. In order to secure such commitments, educational researchers must be prepared to convince school people of the need for and particularly the value of the proposed research. This in turn requires that we give some careful thought to our choice of variables for study, concentrating on those which seem to have the greatest potential for significantly advancing both theoretical understanding and classroom practice.

REFERENCES

Blount, N. S., Klausmeier, H. J., Johnson, Shelby L., Fredrick, W. C., & Ramsay, J. G. The effectiveness of programed materials in English syntax and the relationship of selected variables to the learning of concepts. Technical Report from the Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin, 1967, No. 17.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally and Company, 1963. Pp. 171-246. (Reprinted as Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1967.)

Covington, M. V., Crutchfield, R. S., & Davies, L. B. The productive thinking program, series one: General problem solving. Berkeley, Calif.: Educational Innovation, 1966.

Cox, D. R. Planning of experiments. New York: Wiley, 1958.

Cronbach, L. J. The role of the university in improving education. Phi Delta Kappan, 1966, 47, 539-545.

Ebel, R. L. Some limitations of basic research in education. Phi Delta Kappan, 1967, 48, 81-84.

Hickrod, G. A. Ecological changes within a school district and expenditure for education. American Educational Research Journal, 1967, 4, 241-252.

Lindquist, E. F. Statistical analysis in educational research. Boston: Houghton-Mifflin, 1940.

Olton, R. M., Wardrop, J. L., Covington, M. V., Goodwin, W. L., Crutchfield, R. S., Klausmeier, H. J., & Ronda, Teckla. The development of productive-thinking skills in fifth-grade children. Technical Report from the Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin, 1967, No. 34.

Stanley, J. C. Informal notes about experimental design. Baltimore: Johns Hopkins (Author), 1965.

III

CONTROLLED EXPERIMENTATION IN RESEARCH AND INSTRUCTION UNITS

Mary R. Quilling
Coordinator of the Technical Section
Wisconsin R & D Center

Educational researchers have long been plagued with the problem of finding enough experimental units to conduct research in the school setting. Typically, intact classrooms are available to the researcher, and the classroom is the only possible unit of randomization. In these instances the proper unit of analysis is the mean of each classroom. Of course, a large number of classrooms involving hundreds of children is required to provide data for a sensitive statistical test.

In the event that pupils within classrooms may be randomly assigned to the experimental treatments, a different problem is encountered. Experimental arrangements, such as the process of randomization and splitting into groups, the use of different classrooms and strange teachers, may be so unusual to the pupils involved that these arrangements interact with the treatment. Such reactive arrangements, as they are termed by Campbell and Stanley (1963), jeopardize the generalizability of results.

In response to these experimental design problems, Research and Instruction (R & I) Units were conceived and established in 1966 in elementary schools of five Wisconsin cities. An R & I Unit is staffed by a Unit leader, 3-6 certified teachers, and several paraprofessionals. The Unit leader typically has some advanced training in curriculum and educational psychology, including methodology. This team is responsible for the education of 100 to 200 children, perhaps from two or three grade levels. While distinctive features of this organizational plan incorporate the better aspects of team teaching, let us concentrate on the research capability of these Units.

Of major importance is the flexibility which the Unit organization permits. Teachers and pupils typically change rooms as pupils are regrouped for instruction throughout the day. Assignment of pupils to new groups, instructors, or classrooms thus is an ordinary rather than unusual experience. Experimental arrangements requiring random assignment of pupils to groups are thus far less likely to react with the treatment than would be the case in experiments involving pupils from self-contained classrooms.

Teachers as well as pupils may be randomly assigned to treatments; furthermore teachers may be rotated among treatments so that a potential source of confounding is eliminated. In the R & I Unit staff, the experimenter has persons to administer the treatments who are both qualified to teach and have some appreciation of requirements of experimental rigor. Use of teachers as experimenters not only makes the research generalizable to usual school situations where children are taught by certified teachers; it also exposes the teachers to new methods, contributing to their professional growth. Thus we may conclude that the R & I Unit provides not only a propitious setting for the conduct of educational research, in which several shortcomings of other arrangements are overcome, it has some beneficial effects as well.

This is not to say that all the problems of conducting a true experiment in the school setting are solved. Strictly speaking, performing the randomization on pupils is a necessary but not sufficient condition for their being treated as the unit of analysis. If the pupils are instructed as a group after randomization, then intrasession history, as well as the treatment, can affect the measurements gathered on a particular group. In several R & I Units last year, however, the experimental treatments were individually applied, and thus even this requirement for a true experiment was met.

Let us now consider two of the many experiments which were conducted last year in R & I Units. The experimental treatments in each instance were individually applied after pupils were randomly assigned to treatments. Both experiments represent attempts to translate motivational theory into practice in the school setting. The significance level for testing each null hypothesis was set prior to the experiment at .10, a level considered appropriate for preliminary research.

The first experiment was concerned with effective ways of implementing the principle that pupils should receive informative feedback relative to their school performance (Klausmeier, Quilling, & Wardrop, 1968). Conducted in an inner-city school in Racine, Wisconsin, the experiment was designed to assess the effect of feedback on acquisition of mathematics concepts. The subjects were 72 children in a second-grade R & I Unit classified as "disadvantaged" under Title I terms. The IQs on the Kuhlmann-Anderson ranged from 72 to 136, with a mean of 101. In October of second grade, the mean grade equivalent score on the arithmetic subtest of the Stanford Achievement Test, Primary I battery, was 1.6. Pupils were stratified by age—younger or older—and by sex. Twenty-four children, six from each age—sex group were randomly selected as experimental subjects. The remainder of the group served as a control.

All pupils received individualized mathematics instruction. To assist teachers in assessing each individual's progress, an individual record folder was designed. All the major concepts and subconcepts in mathematics were identified and compiled into a type of checklist in the individual folder. Beside each concept was a square to be colored in as an indication of the mastery of that concept. Diagnostic tests were developed for each main concept and subconcept. The format of the folder made it appropriate for use in parent-teacher and pupil–teacher conferences.

After each child's initial standing was determined using diagnostic tests and teacher observations, instruction was prescribed at an appropriate level. As many as four different instructional groups, each focusing on a different concept, were conducted simultaneously. Children were continually shifted among groups as concepts were mastered and new skills required. When a child appeared to have persistent difficulty in grasping one concept over a long period of time, he was allowed to go on to another group and return to the troublesome concept later. In this way, no child met consistent frustration, and individual mobility was assured.

The experimental pupils additionally met individually with a teacher for five minutes each week to discuss progress and identify goals for the following week. At this time the child was allowed to color in the square for the concepts acquired during the week and was praised and encouraged by the teacher. Each teacher met individually with six of the experimental students every week for 5 weeks; then teachers were rotated. In this way, the effect of a particular teacher on a child's progress was minimized. The duration of the experiment was 20 weeks. Each pupil, experimental or control, had access to his folder at all times.

Two dependent variables were used to assess the effects of the experimental treatment — the number of concepts mastered (or squares colored in) during the experiment and the score obtained on a teacher-constructed posttest which sampled the tasks enumerated in the mathematics folder. The means for each group are presented in Table 1.

Table 1

Means on Dependent Variables for Assessing the Effect of Feedback on Acquisition of Mathematics Concepts

	Group	
	Experimental	Control
Concepts mastered	100	75.0
Score on posttest	79	63.5

A multivariate analysis of variance was performed on the data. The results indicated that the treatment was significant at the .10 level.

Since the superior performance of the experimental group can be attributed to social interaction, goal setting behavior, and praise as well as to informative feedback, experiments have been designed this year to clarify the contributions of specific components of the treatment. In summary, the preliminary experimentation identified a motivational procedure which was successful in the classroom setting; thus encouraged, the teaching staff is ready to conduct more rigorous experiments which will further contribute to knowledge of the school application of theoretical constructs. Involvement of school personnel in generating the questions and executing the treatments has, in this instance, made possible the iterative experimentation so necessary to learn why a procedure works. In an iterative cycle of experimentation, the results of an initial

experiment are used to plan a subsequent refined experiment in which further information may be gained.

A second experiment was planned to learn whether older children are effective models for younger children of the same ethnic background (Quilling, Cook, Wardrop, Klausmeier, Baldwin, & Loose, 1968). Subjects in this experiment were educationally disadvantaged children in a Milwaukee inner-city school. The majority of subjects were Negro and handicapped by such characteristics as short attention span, poor self-concept, inadequate motor skills, and low IQ. The Pintner-Cunningham Test administered in the fall to the first graders revealed a mean IQ of 81 with the range being from 55 to 105.

Fifty-seven children in the first-grade R & I Unit were involved in the experiment. From these, 22 (11 boys, 11 girls) were randomly selected as experimental subjects to work with models. The remaining 35 served as the control group. Similarly, 22 sixth graders (11 boys, 11 girls) were selected from the group of sixth graders nominated by their teachers to be appropriate models. The older pupils came from the same physical and socioeconomic environment as the primary children.

Each sixth-grade "helper" was randomly assigned to an experimental subject of the same sex. The older-younger pairs so formed spent one-half hour per week together playing arithmetic games, manipulating concrete objects, and helping with problem solving in an Instructional Materials Center environment. In order to minimize the disruption of schedules, the models reported at times most convenient to their individual timetables. Additionally, the sixth graders were briefed for half an hour each week on appropriate activities for the younger children and reinforcement procedures. The three primary teachers each took responsibility for supervising the experimental treatment on a rotating basis.

All primary children received the same mathematics instruction four days a week. A televised mathematics program, Patterns in Arithmetic for Grade 1, was used together with an accompanying manual and teacher-made worksheets. Large and small group instruction was used as appropriate. In addition, individual instruction was given to pupils when considered necessary.

A teacher-constructed test was designed to measure acquisition of the mathematics concepts presented during the 15-week experiment. The test was administered to all first graders at the conclusions of the experiment.

The means of experimental boys and experimental girls were higher than those of control groups of each sex, as Table 2 indicates.

Table 2

Means on Teacher-Constructed Test to Assess the Effect of Models on Learning Mathematics Concepts

Group	Boys	Girls	Mean
Experimental	16.00	21.27	18.63
Control	13.78	17.84	15.80

The analysis of variance performed on the scores indicated that both treatment and sex were significant sources of variation. The statistical analysis is presented in Table 3.

Table 3

Analysis of Variance on Scores of Teacher-Constructed Test

Source	df	MS	F
Treatments (T)	1	108.0886	3.3822 ($p < .0716$)
Sex (S)	1	295.9425	9.2604 ($p < .0037$)
TS	1	5.0827	.1590
Error	53	31.9015	

Evaluation of the experiment also indicated that all experimental pairs were not compatible. Thus experiments in which the pairings are maximized for compatibility and model-effectiveness are a logical outgrowth of the one reported here.

Experience to date indicates that given significant results, the teaching staffs are eager to incorporate the treatment into the instructional program or to investigate further the effectiveness of the procedure. The interplay of ideas of the classroom teacher and R & D specialist—be he an educational psychologist, a curriculum expert, or a research methodologist—have resulted in empirical evidence of the applicability in the school setting of principles derived from motivational theory. That instruction has been improved simultaneously is evident from the field testing results of schools with R & I Units (Klausmeier et al., 1968; Quilling et al., 1968).

In summary, the R & I Unit is so organized that the conduct of true experiments in the school setting is possible. Staffing and physical arrangements allow random assignment and individual treatment of pupils. Exemplars of such experiments, in which motivational theory was translated into practice, were given. Of promise for the future is the climate for iterative experimentation which has

been created. Already results from initial experiments have been used to design subsequent experiments by which teaching-learning theory may be further refined.

REFERENCES

Campbell, D. T., & Stanley, J. C., Experimental and quasi-experimental designs for research on teaching. In Gage, N. L. (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Klausmeier, H. J., Quilling, M., & Wardrop, J. L. (Eds.) Research and development activities in R & I units in five elementary schools of Racine, Wisconsin. Technical Report from the Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin, 1968, No. 52.

Quilling, M. R., Cook, D. M., Wardrop, J. L., Klausmeier, H. J., Baldwin, R., & Loose, C. Research and development activities in R & I units of two elementary schools of Milwaukee, Wisconsin. Technical Report from the Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin, 1968, No. 46.

IV

THE DEVELOPMENT AND REFINEMENT OF PROTOTYPIC INSTRUCTIONAL SYSTEMS

Thomas A. Romberg
Assistant Professor of Curriculum and Instruction
The University of Wisconsin
Director of Program 2, Processes and Programs of Instruction
Wisconsin R & D Center

The purpose of this paper is to describe a major activity of a number of projects within Program 2 of the R & D Center, namely the development and refinement of prototypic instructional systems.

Program 2, Processes and Programs of Instruction, is one of three major programs of the Center attempting to improve educational practice utilizing knowledge about cognitive processes and applying this knowledge to instructional problems. The program has five general objectives:

1) To establish rationale and strategy for developing instructional systems in the cognitive domain.
2) To identify by careful synthesis and further research sequences of concepts and cognitive skills within and across disciplines.
3) To develop assessment procedures and materials for the concepts and skills identified above.
4) To identify existing materials or develop new instructional materials associated with the concepts and cognitive skills.
5) To generate new knowledge through research about instructional procedures including motivation, individualization, classroom management, and organization of instruction.

In order to reach these objectives, instructional programs in mathematics, English, reading, speech, and science are currently being studied. One aspect of this overall effort is the development of prototypic systems.

The development of prototypic instructional systems involves an attempt to spell out and validate operational plans and specifications for instructional systems which have certain components. Prototypic, stemming directly from prototype, implies that what is to be developed is intended to be not a fully developed program, but guidelines or specifications for complete programs. The word was chosen to convey the fact that we are neither developing complete instructional programs or materials nor competing with large curriculum projects. A prototypic instructional system is only intended to be a framework for further development.

The strategy being followed is quite different from that used by curriculum projects such as SMSG or PSSC. None of those curriculum projects attempted to change in any way the basic educational units of our school system. They only attempted to change the content of physics or mathematics or chemistry within the existing school organization. For example, algebra has been traditionally taught one hour a day five days a week in groups of 30 with one teacher. Developers of new programs made no attempt to change that pattern; instead, they worked within the existing structure of most American schools and restricted the revision to subject matter.

However, some schools today are now, and most schools of some not too distant tomorrow will be, quite different from the stereotype school those curricula were developed for. With the advent of flexibly organized schools, such as the Multiunit School, more carefully engineered instructional programs must be invented.

One common basic function of contemporary elementary schools is "to execute a system-wide standard instructional program designed by others [Klausmeier, Morrow, & Walter, 1968, p. 1]." On the other hand the school of the future will be "developing and executing

an effective system of individually guided education within each building [Klausmeier et al., 1968, p. 1]." Thus, the long unrealized goal of providing instruction that is effectively geared to the individual learner is now possible. The intent of developing prototypic instructional systems is to construct the specifications for individually guided instructional programs.

The strategy being used has been (1) to identify component attributes of an instructional system which might be manipulated or changed, and (2) to construct sample elements for an instructional system which can operate within a flexible educational system.

The word system as used in the paper refers to a man-made controlled functional structure. Man-made structure means that the system has interdependent components which can be changed or manipulated. Controlled means that there is a feedback or monitoring procedure which can be used to manage the system, and functional means that the system is goal oriented with a stated purpose or intent. Minimally such a system has four basic components: input, mechanism, feedback and output. The input is raw materials; the mechanism is the way in which the raw materials are transformed into the output; the feedback is the control process; and the output is some product reflecting the purpose of the system.

In an instructional system the input is students; the mechanism is the instructional program; the feedback is an evaluation program; the output is the same students with changed behaviors. There is also, in an instructional system, a resource component which includes instructional materials and equipment, teachers and staff, and facilities and their allocation. Each component could be viewed as a subsystem with internal, interdependent elements. Three elements of the evaluation component—monitoring procedures for the input, the output and the resources of the system; a data collection and management element; and decision-making procedures in which criteria and explicit decision rules are established—are identified in Figure 4, and their relationship to each other and to other components is illustrated. Thus, the development of prototypic instructional systems involves detailing the design specifications for each component and the relationships of the components.

In developing operational plans for an instructional system, activities are carried out to answer a number of queries. Why? requires the identification of purposes and goals. What? is answered by an outline of content. To whom? necessitates consideration of performance history, organismic variables, and cultural influences. When? demands specification of time allocation and sequencing of the content elements. Specification of transmission procedures answers How? Does it work? is answered by the development of

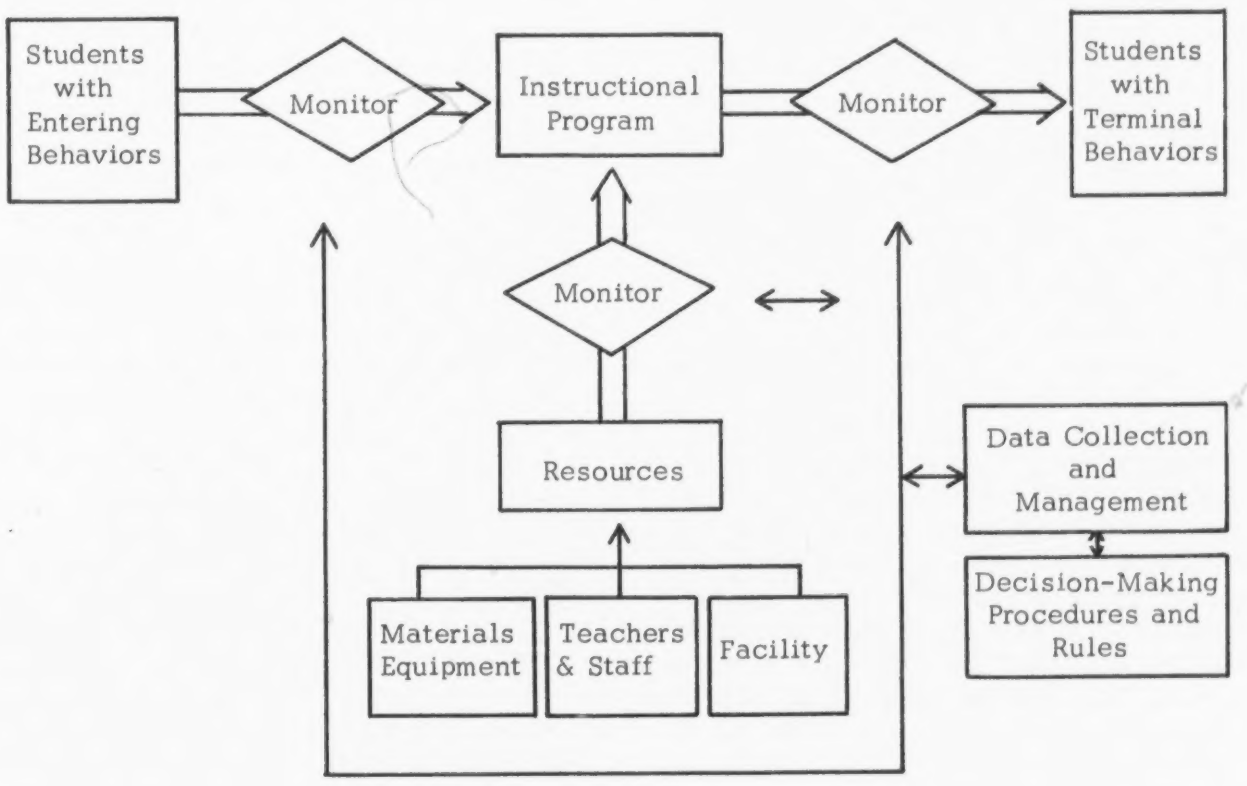

Figure 4. Model of an Instructional System

measurement tools and of evaluation and decision-making procedures.

Goals are statements of intent which give meaning to content selection. They may be broadly conceived for the whole system, as utilitarian or cultural goals whose establishment is a societal responsibility. Or goals may be more narrowly conceived in terms of basic intellectual disciplines. Separate instructional systems can be developed for reading, mathematics, science, etc. as subsystems of a total educational system. This second type of goal is the focus of Program 2 of the Center for disciplines with high cognitive content and agreed upon cultural value. Prototypic instructional programs for the elementary school are currently being developed in reading, science, and mathematics. Similar programs in language arts and speech are being worked on at higher grades.

Initial conceptualization of a content-based instructional system is done by outlining the content to include concepts (facts, assumptions, definitions, symbols, etc.), principles (associations between facts), conceptual schemata or structure (the organization of a set of facts and principles), and methods (the ways in which scholars use facts, principles, conceptual schemata, and skills). But an outline is not sufficient. Descriptions of observable behavioral characteristics for students who have acquired each concept or mastered each skill must be prepared. The translation of content into behavioral objectives or desired terminal behaviors is critical since statements of observable performance give direction to the development of evaluation procedures.

Individual differences between learners are well documented. However, only a small part of this knowledge has been related to instructional programs. The abilities of learners are dependent upon their developmental level, their intellectual power, their learning style, their past learning experience, their parental relationships, their peer group associations, etc. In addition, the emerging picture indicates that performance history (the nature and quality of prior experiences) in a subject area is particularly important. One of the critical unsolved engineering problems in the construction of individualized instructional programs is how one can incorporate this wealth of information about individual differences into the system so it can be used to guide an individual's learning experiences. This information needs to be included as one considers each step in the development of operational plans.

Once the content domain has been specified, the next task is to begin to put the concepts and cognitive skills together in some sequence and to allocate time, age level, number of days, etc. which will be assigned to this concept or that skill. This is not an easy task. It involves determining prerequisite behaviors or prerequisite learned concepts and skills. For each skill it involves examining pupils and their general cognitive development, and it involves consideration of alternative sequencing based on performance histories. Clearly there is no one best sequence. Alternative sequences and time allocations must be developed.

Given an outline of content and cognitive skills and a sequence and time allocation of these skills for an instructional program, the details of an instructional program can be described by a list of the ways in which these concepts and skills can be transmitted to students. This involves two steps: first, describing potential student activities which, if followed, reach desired objectives; and second, specifying what resources are to be used and how they are to be used. The focus of how information is to be transmitted to students has to be on the activities he is asked to perform. The character of the activity is dependent upon both what is to be transmitted and the personality of the student. Again alternative activities must be invented.

Resource allocation involves detailing the teacher's activities in organizing and directing the student activities. It also involves specifying what instructional materials and/or supplies are to be used. And, it further means that the organization for instruction—i.e., nongraded, self-contained, unit or team; one subject, more than one subject; modular time schedule, etc.—be detailed.

Concurrently, measurement tools and evaluation procedures need to be developed to measure students' prior achievements or readiness to engage in the program, to measure the students' more general aptitudes and characteristics including their social background characteristics, to measure progress during short and long intervals of instruction, and to evaluate the components and total system including student progress and level of achievement. To do all this involves spelling out what is to be monitored with respect to initial behavior, terminal behavior, and resource allocation; how these data are to be collected and managed; and what kinds of decisions are to be made on the basis of these data.

Our long-range plans require consideration of each area mentioned. However, for short-range planning emphases are only on selected components. Projects in reading, science and mathematics are described below to illustrate what is now being done.

GPO 811-034-3

In reading, Professor Wayne Otto and his staff working with school personnel have identified reading skills and assessment procedures (Otto, Saeman, Houston, McMahan, & Wojtal, 1967). The intent was:

> to go beyond the simple sequencing of skills by attempting to suggest specific exercises and/or observations designed for use in assessing mastery or progress in the development of each of the several skills. Concurrently, an attempt is made to avoid limiting the use of the Guide to a particular instructional setup, e.g., ungraded, self-contained, interclass, etc. The specific suggestions for assessing skill development should provide not only a means for gathering information but also a guide to systematic teaching of skills.

Current work in the reading project is to refine and clarify these skills and assessment procedures and to extend the Guide to include a catalogue of materials which can be used to teach each skill.

A second example is the science project headed by Professors M. O. Pella and G. T. O'Hearn. Science, in contrast with reading, has not normally been taught in elementary schools, hence, there is no accepted set of concepts and skills. Identifying these content elements and seeing whether young children can acquire them is a major part of their current efforts. In addition, they are engaged in establishing behavioral objectives and developing related assessment tools.

In mathematics, another approach is being taken by Professor H. J. Fletcher and myself. We began with broad mathematical objectives and then identified the concepts, skills and operations (basic units) involved in reaching these objectives. After the basic units have been compiled, the following steps are taken:

a) Each unit or concept is analyzed in terms of its subconcepts, properties, or attributes, together with any rules necessary for combining them.
b) Prerequisite behaviors the student must possess for any unit are identified.
c) These prerequisite behaviors are used to develop a logical sequencing of the units.
d) Student behaviors are associated with each unit. Observation of these behaviors provides the only method for evaluating change in student behaviors.
e) Teaching tactics are identified for each unit. Tactics include determining whether the student has the necessary prerequisites for a unit, assembling the necessary material, performing demonstrations, posing problems, etc.
f) Problem situations are designed to secure and hold attention, or to maintain high motivation.

The above steps outline the conceptual analysis phase of our project. A pilot examination phase follows and consists primarily of exploratory pilot research in which each problem situation is tried out with a few students in a normal classroom. In a subsequent evaluation phase the effectiveness of each problem situation is assessed. Unfavorable results would necessitate returning to an earlier phase; favorable results permit continuing into the next phase, validation, in which careful studies are conducted on all elements of the system prior to field testing of the entire system.

What these examples imply is that we are trying to conduct <u>research-based development</u>. The problem, although initially an instructional development problem, must be conceptualized in terms of knowledge about learning (concept formation, rule learning, problem solving, motivation, etc.). The knowledge and methods of the science of psychology must be brought to bear on instructional problems. But the connection goes both ways. Many real instructional problems have not been adequately studied within a theoretical or laboratory framework. It is now becoming apparent that much basic research will be <u>development-oriented research</u>.

To construct a prototypic instructional system real interdisciplinary scholar-practitioner teams must be working together on the problem. In the mathematics project mentioned above a psychologist, a mathematician, a mathematics educator, a classroom teacher, and five research assistants (in both mathematics and psychology) are involved. One must also have an appropriate environment in which to try out one's ideas. The Multiunit School developed by Professor Klausmeier and his staff at the R & D Center is proving to be an excellent environment in which to do programatic research. In addition, one must have the luxury of time to do long-range programatic research. The R & D Center concept provides for long-term projects aimed at major conceptual and programatic problems.

In summary, prototypic systems of instruction are operational plans for educational systems which have been carefully formulated and validated. We believe that the development and refinement of prototypic instructional systems as being done at the Wisconsin R & D Center will significantly influence American education in the future.

REFERENCES

Klausmeier, H. J., Morrow, R. G., & Walter, J. E. Individually guided education in the Multiunit elementary school: Guidelines for implementation. Madison: Wisconsin R & D Center, University of Wisconsin, 1968.

Otto, W., Saeman, R., Houston, C., McMahan, B., & Wojtal, P. Prototypic guide to reading skill development in the elementary school. Working Paper from the Wisconsin R & D Center for Cognitive Learning. University of Wisconsin, 1967, No. 7.

V

EXPERIMENTATION RELATING TO FORMATIVE EVALUATION

Richard E. Schutz
Director, Southwest Regional Laboratory

I am going to restrict my remarks to one-half of the title of the symposium. I will be talking about development strategies and instructional improvement and will not attempt to deal with research or theory refinement. Research and theory go together like development and improvement. The distinction is primarily in terms of the outcomes generated. Research produces refined knowledge; development produces usable products. Research seeks the answer to the question How now? Development seeks the answer to the question How to?

Scriven (Tyler, Gagne, & Scriven, 1967), you will recall, introduced the term formative evaluation to describe the evaluation of educational programs that are still in some stage of development. Formative evaluation contrasts with summative evaluation, the evaluation of finally developed educational programs. The product of formative evaluation activities is expected to be an improved instructional program, while the product of summative evaluation is normally a set of descriptive statements about a single program or about the relative merits of two or more programs. An extensive methodology is associated with summative evaluation, since researchers have had extensive experience in comparing two or more existing programs. Formative evaluation, however, is a different matter.

At present, formative evaluation methods have much the same status as the invisible needle and thread used by the tailors in the "Emperor's New Clothes." For a variety of social, political, and economic reasons, educational researchers have derived satisfaction in viewing their efforts as contributing to educational improvement. But the inescapable fact is that reliable ways of effecting educational improvement have yet to be identified either by researchers, manufacturers, or school personnel.

J. M. Stephens, in his highly provocative little book The Process of Schooling (1967), summarizes the summaries of experiments on instruction over the last 50 years. This is a disturbing treatise which I highly recommend. Stephens comments upon "the remarkable constancy of educational results in the face of widely differing deliberate approaches." He goes on to state bluntly, "Every so often we adopt new approaches or new methodologies and place our reliance on new panaceas. At the very least we seem to chorus new slogans. Yet the academic growth within the classroom continues at about the same rate, stubbornly refusing to cooperate with the bright new dicta emanating from the conference room [p. 9]." A spectacularly elegant NSD.

Stephens' prescription is to relax and enjoy the powerful and pervasive educational forces which apparently work well apart from any deliberate direction. Get a few adults, put them in contact with kids for as much time as possible, and count on spontaneous factors to take care of everything else.

I find this line of reasoning a very threatening but compelling argument, things being what they are. The aim of development, however, is to insure that things do not remain as they are. Development seeks to produce materials and methods called products which can be used to produce a specified outcome. That the desired outcome can be produced is always a matter of faith while the development is in progress. There is no guarantee that we will ever place a man on the moon. Neither is there any guarantee that we will ever teach all children to read. But in each example,

there is a high probability that the investment of cumulative effort over time will achieve the desired end.

Educational development is at present in a state of conceptual deprivation. All of the terms available for describing instruction have reference to instructional operation systems—keeping school—rather than to instructional development. For example, grade and subject-matter dimensions and pupil characteristics are useful in defining school operations. To indicate that someone is producing a high school American history program, for example, appears to communicate a good deal. However, the information relates almost exclusively to use of the instruction in the schools. It indicates very little about the instruction per se and indicates nothing about the operations involved in producing the instructional materials and procedures.

The principal dimension traditionally manipulated in instructional development has been subject-matter content- à la the new math, science, and social studies "curriculum development" projects. Formative evaluation efforts have been highly content oriented. Measurement results have been used as a basis for making decisions dealing with the sequencing of substantive concepts and for modifying the instructional verbage associated with a given concept. When the measurement results are not easily interpreted, one of two events takes place. (1) The proposed outcomes of the instruction are modified, or (2) the goals are redefined in more general or abstract terms to facilitate agreement about the interpretations of the measurement efforts. But substitutions in instructional goals should not be confused with improved attainment of these goals. Nor is it defensible to retreat behind a mystique of complicated and intangible criteria.

If one takes a product-oriented approach, the criteria for gauging instructional improvement become straightforward and the application of formative evaluation procedures more meaningful. Remember that we have defined a product as consisting of the materials and accompanying procedures to accomplish specified instructional outcomes. In this setting we are concerned with three criteria dimensions: reliability, utility, and cost.

We must be concerned first with the reliability of the product in accomplishing the objective. How well does it work? Here we are considering such things as dependability of the outcome, generalizability or exportability to a variety of locales, and replicability. Until an effect can be produced with an acceptable degree of dependability, there is really nothing for a potential user to evaluate. Thus reliability of effect is an important criterion.

A second set of criteria involves utility. How useful is the result? Here we can distinguish between social utility, which involves an outcome that is useful at present, and instructional utility which involves an outcome that provides a base for future educational attainment. To be useful, a given outcome must have clearly demonstrable social and/or instructional utility.

When we are able to produce an instructional effect with a given degree of dependability, and it is judged to be worth producing, then we must necessarily be concerned about cost. Here we are talking about either the time or the money required to produce a given result. Any form of cost can be transformed into one of these two dimensions.

It is only when one is able to produce useful outcomes with a reasonable degree of reliability that it makes any sense to talk about cost benefit analysis. But with reliability, utility, and cost targeted you can very reasonably talk about both cost efficiency and cost effectiveness. Cost efficiency involves the ratio of reliability to cost; cost effectiveness involves the ratio of both reliability and utility to cost.

Education has yet to cope adequately with the first set of criteria—reliability of effect. We simply lack the wherewithal at this time to dependably accomplish the attainment of the educational outcomes for which the schools have assumed responsibility. At present we can at best plan strategies which have some likelihood of leading to the improvement desired. Several general alternative strategies can be formulated which themselves will provide the basis for a grand experiment. I shall limit my considerations to the experimentation possible within a single such general strategy.

The strategy is derived by analogy from engineering; underscore analogy. Analogy is the weakest form of argument, but the relationship at present between educational and engineering development is analogous at best.

The foundation of the strategy is planned iteration—a set of sequenced and coordinated activities which through careful management cumulate to effect the improvement desired. That's fine. But, what do we iterate? Or, what should we iterate? Well, we have to start with what we've got. Here we can identify several manipulable dimensions of instruction. For example:

> The instructional media—at present this is typically the teacher or printed matter, although a wide spectrum of auditory and visual presentation mechanisms are possible.
>
> The pacing mode—at present instruction is group paced, although

various degrees of differentiation toward individual pacing can be identified.

The monitoring contingency—at present the attainment of instructional outcomes is norm-referenced, if contigent at all. That is, one lesson typically follows another, irrespective of whether all learners have attained the outcomes the lesson was designed to produce.

The final example of possible dimensions is the monitoring agent. At present, this is the teacher. The teacher is responsible for insuring that the instruction is sequenced, coordinated with other activities. But the pupil himself could make a contribution in the area as could other pupils, parents, computers, etc.

Describing current school practices using the four dimensions of (1) Instructional Media; (2) Pacing Mode; (3) Monitoring Contingency; and (4) Monitoring Agent, formal public education at the present time can be characterized as print mediated, group paced, managed against relative norms, and teacher-based. There is good reason to believe that instruction which is multiple mediated, individually paced, managed against objectives, and computer-based would be much more effective. My hunch is that it will be at least a decade before such a system is operationally feasible. But this hinges on the referents for the term feasible. From a researcher's point of view, any phenomenon which can be conceptualized and potentially manipulated is regarded as feasible. If the development does not occur easily, the researcher honestly believes that the persons reponsible for translating his efforts into usable instructional products and operating the instructional program have hopelessly fouled-up the concepts. The developer, on the other hand, finds himself in an awkward "Mr. In-between" role. He views the researcher's ideas as perpetually incomplete and inadequate as a development guide. At the same time, he necessarily views current school operations as obsolete since his efforts are directed toward improving it. School personnel, for their part, forever find the present work of both the researcher and the developer unrelated to immediate everyday problems. Each is accurate. Moreover, this is not an undesirable state of affairs so long as it is recognized; it provides the basis for an efficient division of professional labor.

How is this educational improvement to be effected? I should like to suggest two concurrent and compatible tactical approaches. Each recognizes the present print-mediated, group paced, weakly managed, and teacher-based instructional system. One approach involves a series of straightforward trial-revision cycles to sharpen up the accomplishment of given instructional outcomes, working within the boundaries of the present instructional system. An analogy is provided by the modifications in the Volkswagen annually to improve its performance in specified ways. The second approach involves a major manipulation of a system dimension per se which produces a new generation of the product. Following through on the automobile analogy, this might be an electric car.

Each of these two approaches involves a convergent iterative methodology. Let me mention the sequence we find useful at the Southwest Regional Laboratory in pursuing the first approach. It involves first the preparation of instructional specifications. This is a set of sequenced statements of desired instructional outcomes accompanied by specifications of requisite entry behavior, learner activities, specific conditions under which practice must be given to assure that appropriate behavior does and does not occur, and specifications for testing if the outcome is attained.

Component preparation and tryout involves the initial production and tryout of component methods and materials which relate to various aspects of the instructional specifications. These tryouts involve "mock-ups" designed to reflect the critical features of the aspect of instruction about which further information is needed before it can be considered a reliable prototype component.

Product preparation and tryout involves combining and extending the prototype components into instruction suitable for classroom use. The objective is to produce methods and materials which are attractive to pupils, manageable by teachers, at as low a cost and requir- as little instructional time as possible. The product tryout is conducted by regular school personnel under standard school conditions.

Each of these stages involves a continuous sequence of trial-revision iterations to successively eliminate the defects in the product and to increase its effectiveness. The sequenced instructional objectives provide the criterion against which base-rate performance data can be generated. This provides an assurance that the changes introduced cumulate in improvement and avoids expensive unevaluated modifications which fail to improve overall performance.

The principles of scientific experimentation —for example those discussed by Campbell and Stanley (1963) under the rubrics internal and external validity—are highly relevant at

each stage of the cycle. But, development is concerned with the management of variance rather than the analysis of variance. The comparisons of interest involve variations and refinements introduced sequentially over time rather than concurrently at a single time. Note that I'm not talking here of a time-series design in the Campbell-Stanley sense. Treatments A_1, A_2, etc. are compared sequentially in toto as each treatment is successively reviewed, refined, and recycled on fresh samples. Treatment A_2 always incorporates "the best" of A_1; A_3 the best of A_2; and so forth. If a given iteration yields worse rather than better results, one goes back to the precondition and reformulates the modification to be made. Our experience to date has been that this iterative approach does pay off.

So much for the first approach, effecting cycles of improvement. In effecting generations of improvement one is seeking to introduce a modification in at least one system dimension with the objective of concomitantly increasing effectiveness along other dimensions. This involves complex development work that is completely removed from the refinement cycles within a present generation.

Consider the instructional monitoring dimensions. The Southwest Regional Laboratory, in conjunction with System Development Corporation, is currently engaged in the development of a computer-based instructional management system. Here the immediate concern is not with all instructional functions, but simply in developing workable procedures to provide the teacher with frequent information concerning the progress of each student with respect to the instructional outcomes of interest to the teacher together with suggestions for activities which are consistent with the student's performance.

Thus in contrast to computer-assisted instruction system, we are seeking a technologically less sophisticated but more feasible computer-managed instruction system. The computer technology problems are being solved using the same iterative strategy outline above. However, rather than pupil performance, the criteria involve teacher management performance. Although live classes are being used in developing the system, there is no expectation that pupil performance will be improved at this point. However, when the system is functioning adequately, the greater monitoring capability it will provide is expected to improve performance sufficiently to introduce a new generation of instructional effectiveness. Analogous programatic efforts in media and pacing are also being conducted by the Laboratory.

In sum, educational development appears to be as vast and promising an enterprise as nuclear or space development. Just as engineering enterprises find it necessary to use all existing relevant knowledge plus, so does educational development. With this effort the concept of formative evaluation will be clothed with a technology of instructional product development. This is at once a challenging and threatening enterprise for those who become involved in it. The "produce or perish" imperative associated with educational development is an even tougher task master than the "publish or perish" imperative of academia and the "profit or perish" imperative of business. However, it is an imperative which the education profession can ill afford to ignore, considering the tenor of the times. The frontier is there, with lots of wide open territory to be pioneered. If you are bright and brave, blast into it.

REFERENCES

Campbell, Donald T., and Stanley, Julian C., "Experimental and Quasi-experimental Design for Research on Teaching" in Gage, N.L. editor, Handbook of Research on Teaching. Chicago: Rand McNally & Company, 1963.

Stephens, J. M., The Process of Schooling. New York: Holt, Rinehart & Winston, 1967.

Tyler, Ralph, W., Gagne, Robert M., and Scriven, Michael. Perspectives of Curriculum Evaluation. AERA Monograph Series on Curriculum Evaluation, Monograph No. 1. Chicago: Rand McNally, 1967.

VI

REMARKS FOLLOWING THE PRESENTATION OF PAPERS

Leslie D. McLean
Associate Professor and Chairman, Department of Computer Applications
The Ontario Institute for Studies in Education

The topic of this symposium, as stated by Professor Klausmeier in his first sentence, is an important and vital one. The papers deal with a specification of the question, How do we organize for research and development in education? It is refreshing to find the presentation and discussion of efforts which have already been tried as opposed to suggestions of what might possibly work. It is probably no accident that the papers are organized along the lines of a research project report. Klausmeier and Romberg provide the theory and background of the problem, Wardrop and Quilling the procedures and results, and Schutz the discussion and critique. It has been a most stimulating and worthwhile symposium and I am pleased and honored to be a part of it.

My first reaction is that we are seeing a new stage in the evolution of experimental design. First there was the movement of design concepts into educational research textbooks. Their actual use by experimenters has been very limited; until recently status studies rather than the controlled experiment were a mark of educational research. The change seems all the less surprising when we look at the magnitude of the effort recorded by the research and development centers. It is clear that a sustained series of meaningful experiments carried on in school settings does require the trained manpower and very considerable resources which the R & D centers can provide. The continuity in an R & D center can provide the support for an iterative program of experimentation where insights and hypotheses discovered in one experiment can be followed up soon enough to get answers and move on to the new hypotheses so generated. The same mistakes need not be repeated quite so often.

It is interesting to compare the R & D centers with the Ontario Institute for Studies in Education. The OISE was founded as an independent college by the Province of Ontario with the task of carrying on research, development, and graduate instruction. Not only is the teaching dimension added, but certain aspects of development and dissemination which have been assigned to the regional laboratories in the United States. Perhaps you can see that the topic of this symposium has direct relevance to the problems being encountered in Ontario.

I am pleased to have reinforced a vague hypothesis which I formulated recently, namely that not only does one follow the path from research to development to dissemination, but development also produces research ideas and pushes one straight back to the important questions. As we make serious efforts to develop worthwhile instructional products, we are constantly confronted with unsolved research questions and constantly challenged to think more clearly about exactly what we are trying to do. Schutz' paper is a most valuable contribution since few places are actually producing instructional products in an atmosphere of research and development.

The phrase "systems approach to education" has been accompanied by so much talk and so little action that I have developed a strong avoidance response to it. I must admit that this response is near extinction after listening to Wardrop's and Quilling's papers, followed by Romberg's paper. Not only are there systems on paper, but there are actually things called Research and Instruction Units. People are doing research within an overall framework with some attempt at coordination. I must again point to the OISE as an example of a

similar attempt; we should be able to increase the probability of our success given useful inputs such as these.

My final comment stems from a general impression I have that everything is arranged and prescribed for the student. The language we use constantly reinforces the idea that "we" know what is good for "them." My feeling is that this is not so much so as we might think it is so, but that we should make a concerted attempt to develop an attitude that the student must begin to exercise more control over his own education. If I am not mistaken, Schutz was the only one to actually mention students and their place in the system. Complete control to achieve specified learning outcome will be quite a failure if the student develops no initiative of his own. Let us by all means carefully consider what we will call success.

GPO 811-034-2